Proceedings of the International Conference on Industrial Engineering and Operations Management
Bangkok, Thailand, March 5-7, 2019

Feature Selection and Machine Learning Classification for Live P2P Traffic

Haitham A. Jamil
University of Elimam Elmahdi
Kosti, Sudan
haithamjamil@mahdi.edu.sd

Bushra M.Ali
University of Technology Malaysia
Johor, Malaysia
Bushra0912115@gmail.com

Entisar H. Khalifa
University of North Border
Arar, Saudi Arabia
ent_osman@yahoo.com

Abstract

Classification of bandwidth-heavy Internet traffic is important for network administrators to throttle network of heavy bandwidth traffic applications. Statistical methods have been previously proposed as promising method to identify Internet traffic based on their statistical features. The selection of statistical features still plays an important role in accurate and timely classification although most feature selection algorithms consider the correlation between features. In this work, we propose a technique based on features characters and Principal Components Analysis (PCA) feature selection algorithms for online Peer-to-Peer (P2P) traffic detection. Using Naïve Bayes and J48 machine learning techniques for available traces from University of Brescia and University of Cambridge, experimental results show that the proposed method is able to achieve up to 99.5% accuracy for 0.007 second testing time. These results are superior to other existing approaches in term of accuracy and testing time.

Keywords
Online features, P2P, features selection, machine learning, Traffic classification.

1. Introduction

Improvements in computing and communication nowadays bring to society tremendous benefits. Today, peer-to-peer (P2P) is considered as a standard architecture for sharing a wide range of various medias throughout the Internet. The high volume of P2P traffic is due to file sharing, video streaming, online gaming and other activities that client-server architecture cannot accomplish as fast or as efficient as the P2P architecture. P2P traffic represents 27% to 55% of the aggregate internet traffic, depending on geographic location [1]. The tremendous amount of P2P traffic and its rapid progression throughout the years have resulted in deteriorated network performance and congestion due to the high bandwidth consumption of P2P applications[2]. Therefore, traffic identification is required to improve the network performance.
First generations of P2P application traffic were relatively easy to be identified due to the use of port-based technique. However, current P2P applications are able to circumvent port-based identification by using dynamic port numbers or port masquerading. Besides, methods that rely on inspecting application payload signatures have also been proposed [3, 4]. Nevertheless, more P2P protocols support payload encryption. The frequency of which P2P protocols are introduced and/or improved renders packet payload analysis is not only impractical but also inefficient. Moreover, some privacy laws may not allow administrators to inspect traffic payload. The diminished effectiveness of the port-based and payload-based techniques motivates the use of flow statistics for traffic identification. These

techniques offer flexibility to detect P2P traffic compared to using signature-based method.
Presently, extensive researches have been proposed over the last two decades that focuses on the achievable accuracy of different Machine Learning (ML) algorithms. However, the impact of using different sets of statistical features has not been researched in-depth. Work in [5] found that feature selection has a positive impact to improve the performance than the selection of the classification algorithm. Presently, several feature selections algorithms have been introduced [6-8]. However, not all of the selected features can be extracted online since some features cannot be calculated before the flow is finished [6].
This paper proposes an approach based on the characterization of features and feature selection in order to provide optimal features for on-line P2P traffic detection. The objectives of this paper as follows. Firstly, to investigate the impact of the integration of online features and inter-arrival time (IAT). Secondly, to determine the optimal features for on-line P2P traffic classification. Lastly, to evaluate the effectiveness of the proposed features practically in terms of accuracy, recall and testing time.
The remainder of this paper is organized as follows. Section 2 introduces ML concepts, feature selection and related works. Section 3 describes the methodology. The evaluation methodology and experimental setup are discussed in Section 4. The results and discussion are given in Section 5. We conclude the work in Section 6.

2. Related Works

Machine Learning (ML) is one of the modern application classification techniques. It has been known as a collection of powerful techniques for data mining and knowledge discovery [9]. ML can be done with two main methods, unsupervised and supervised learning methods. Unsupervised learning essentially clusters flows with similar characteristics together [10]. Supervised learning requires training data to be labeled in advance and produces a model that fits the training data [11].
For these methods, classifiers are applied to learn patterns to classify unknown files. A classifier is a rule-set that is learnt from a given training set. The first work using this technique was introduced by [12]. This approach needs data mining performed in three steps, extracting the features, selection of feature and generating classifier [13].
The aim of feature selection is to select subset features from the input which can efficiently describe the input data while reducing effects from noise or irrelevant features and still provide good prediction of results [8]. Traffic classification can be improved in terms of accuracy and computational performance by using the most relevant features [24].
Numerous research works have been done in the field of feature selection for traffic classification. Work in [14] suggests 249 flow statistical features that can be potentially used in ML traffic classification. However most of them can only be extracted offline. Offline features such as maximum bytes in packet, minimum bytes in packet, and median bytes in packet only can be extracted after receiving complete flows. Work in [15] used all 249 features suggested in [14] derived from packet streams consisting of one or more packet headers. Most of these features cannot be extracted on-line from live traffic for online traffic classification.
Moore in [16] applied Fast Correlation Based Filter (FCBF) feature selection method for feature reduction and Nave Bayes algorithm to assess effects of the feature reduction. The result of the overall classification accuracy based on the sub sets features is 84.06%, which is the best than using all features. Work in [17] applied two features subsets to provide a classified traffic. This work uses flow features subsets on Support Vector Machine (SVM). The classifier accuracy is 70% while the training time is reported at 40 seconds. Work in [18] identified P2P traffic by using SVM and applies random search algorithm for features reduction. However, this work did not include UDP traffic although P2P traffic consists of both TCP and UDP packets.
Online features techniques were proposed in [7, 8]. These works used ten Cambridge datasets and Naïve Bayes to evaluate two feature selection algorithms named Bias Coefficient Results (BFS) and selected online feature (SOF). These works achieved accuracy 90.92% and 93.20%, respectively. Besides, the works considered inter-arrival time (IAT) as one of the proposed online features.
Most researches focus on investigation on the effect of features selection method applied on 249 flow statistical features as suggested in [14] or applied on online features with IAT as suggested in [6, 8, 19]. However, the impact of the integration of online features and inter-arrival time (IAT) with optimal online features selection for P2P internet traffic has not been researched in deep.

3. Overview of the Methods

In this study, we propose an approach based on features characterization and feature selection algorithms to select online features for online Peer-to-Peer (P2P) Internet traffic classification. Since we focus on online flow-based

classification, this paper does not involve the privacy concerns in processing payload of a packet. Using supervised learning technique with 2-class classification, two classes of dataset samples such as P2P and nP2P are required for the learning and classification parts. The dataset samples that are used as features in the training set are correctly labeled and kept in the corresponding P2P and nP2P set (Figure 3.1).

3.1 Online features

Moore et al. in [14] suggests 249 flow statistical features that can be potentially used in ML traffic classification but most of them can only be extracted offline. Features such as minimum bytes in packet, maximum bytes in packet, and median bytes in packet only can be extracted after receiving complete flows. Considering this type of features in classification will delay traffic classification decision until the end of each flow. Hence, this approach cannot be used to classify live network traffic. Based on references [6, 8, 19], we can categorize the types of online features as display in Table 3.1.

TABLE 3.1 Online features description

No	Name	Short Description
1	Destination port number	DP
2	Source port number	SP
3	Packet inter-arrival time	P_IAT
4	Bytes in Ethernet packet	B_Eth
5	Bytes in IP packet	B_IP
6	Control bytes in packet	C_bp
7	Bytes in Ethernet packet (uplink)	UB_Eth
8	Bytes in Ethernet packet (downlink)	DB_th
9	Bytes in IP packet (uplink)	UB_IP
10	Bytes in IP packet (downlink)	DB_IP
11	Control bytes in packet (uplink)	UC_bp
12	Control bytes in packet (downlink)	DC_bp
13	Packet inter-arrival time(uplink)	UP_IAT
14	Packet inter-arrival time (downlink)	DP_IAT
15	Protocol	Application class

3.2 Inter-arrival time (IAT)

Inter-arrival time (IAT) as shown in Table 3.1 is a part of online features. Inter-arrival time needs pre-processing such as normalization in order to become significant. In addition, the integration of packet size and IAT morphing can heavily thwart the classifier [20]. This is because IAT morphing usually involves alternation on direction pattern [20] and depend on different network locations [21]. Therefore, IAT online features are excluded in this work.
Time-related features do not help to distinguish among applications [10, 22, 23]
Creating a statistical signature of an application solely on the inter-packet time is a challenging task due to the time required by an application to generate and transfer packets to the transport layer is masked by the fact that additional time is added due to the network conditions and the TCP layer [20].
One-way-Delay (OWD) measurement, timestamps of the same packet at different network locations is challenging task due to packet similarities [21].
We also perform experiments to study the impact of the integration of online features and IAT. The result explains IAT does not improved accuracy. Therefore, we do not consider IAT online features as part of our proposed features.

3.3 Existing feature selection method

A good selection method for feature subsets for sample classification is needed in order to improve prediction accuracy, and to avoid incomprehensibility due to the large number of features investigated. The feature selection methods used in this work are Chi-squared [24], Information Gaint (IG) [25], GainRatio (GR) [26], Fuzzy Rough Sub set Eval (FRSE) [27], Filtered Attribute Eval (FAE) and Principal Components Analysis (PCA) algorithms [28, 29].

3.4 Machine learning algorithms

The algorithms are considered in such a way that can help to choose a proper model for an on-line Peer-to-Peer Internet traffic classification. The classifiers used in this work were the HoeffdingTree, J48, SMO, Support Vector Machine SVM, and Naïve Bayes (NB) classifiers.

3.5 The proposed method

In this paper, the proposed method selects important feature set that can be used for online P2P internet traffic classification to improve the classification performance in term of accuracy, recall, incorrectly classified instances (ICI) and testing time. In order to propose online statistical features to be used in P2P traffic identification. Firstly, consider online features as suggested in [6, 8, 19] (see Table 3.1) which were selected from Moore's 249 features. Secondly, study the impact of integration of online features and IAT. Thirdly, subset of feature is created based on Chi-squared, IG, GR, FRSE, FAE and PCA feature selection algorithms. These algorithms are used to select positive features. Then, features subsets are evaluated with five ML algorithms are Naïve Bayes, J48, Lib SVM, HoeffdingTree, and SMO. Lastly, the best combination of subset of features and ML algorithm are selected for on-line P2P traffic identification. Figure 3.1 describes features selection phase and classification phase.

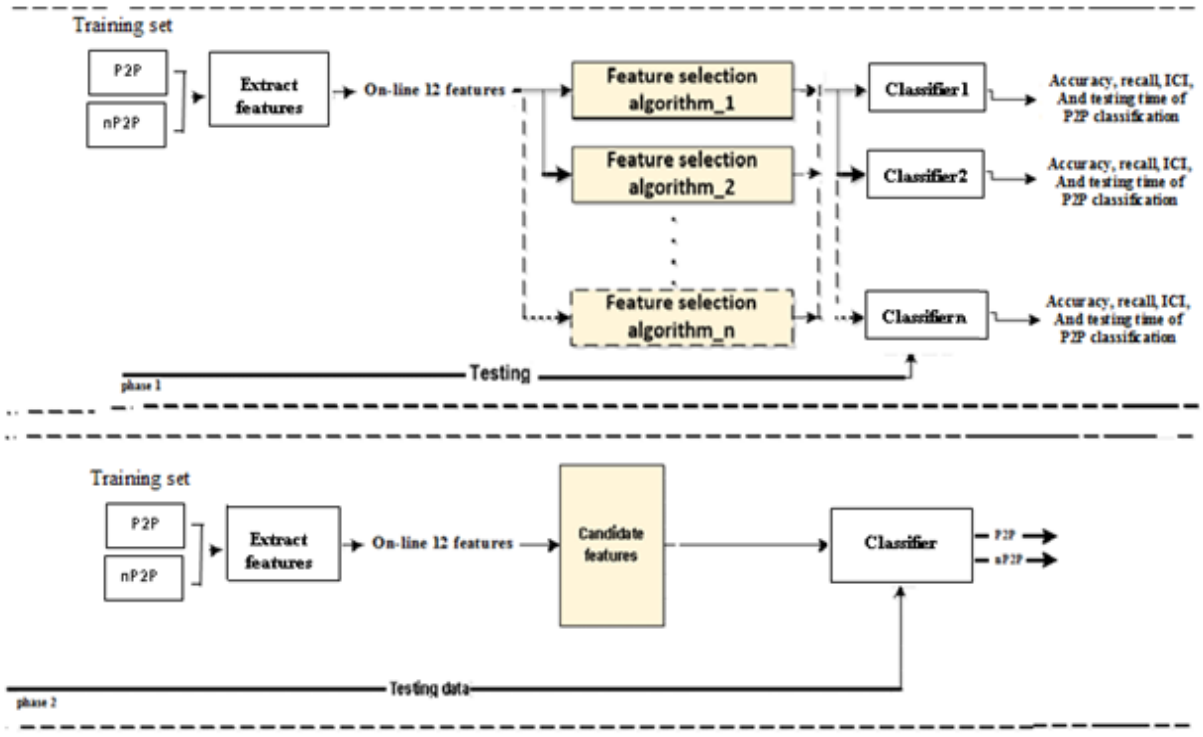

Figure 3.1 Extracting on-line features flow chart

4. Evaluation Methodology and Experimental Setup

In this section, we explain the traffic traces and the evaluation method to evaluate the proposed approach for the selection of on-line features for P2P internet traffic classification. Feature selection algorithms and ML algorithms were implemented using WEKA tool. (WEKA is a collection of open source state-of-the-art machine learning algorithms and data preprocessing tools).

4.1 Dataset

UNIBS traces [33] include packets generated by a series of workstations, located at the University of Brescia (UNIBS) in Italy in September and October 2009.The traces occupy around 2.7 GB (78998 flows) which includes Web, Mail, P2P traffic and other protocols. The details of the UNIBS datasets are summarized in Table 4.1.

Table 4.1 The traces of UNIBS datasets

Dataset	Size
unibs20091002.anon	1.94 GB
unibs20090930.anon	317 MB
unibs20090930.anon	317 MB

Cambridge datasets are based on the traces captured on the Genome Campus network in August 2003. They are published by the computer laboratory in the University of Cambridge[14]. There are ten different datasets each from a different period of the 24-hour day. The number of flows in each dataset is different. These datasets consist of TCP flow. Moreover, each flow example is high dimensional since it consists of 248 features that are derived from the TCP headers by using TCP trace, however, processed dataset of all features are offline. The details of the Cambridge datasets are summarized in Table 4.2.

Table 4.2 The samples of the Cambridge datasets

Dataset	Instances	Size
Dataset1	24863 flows	29.7MB
Dataset2	23801 flows	28.3MB
Dataset3	22932 flows	27.5MB
Dataset4	22285 flows	26.6MB
Dataset5	21648 flows	25.8MB
Dataset6	19384 flows	23.1MB
Dataset7	55835 flows	66.0MB
Dataset8	55494 flows	65.6MB
Dataset9	66248 flows	78.3MB
Dataset10	65036 flows	77.1MB

4.2 Evaluation metric and validation

Performance and accuracy of the proposed method are validated by using incorrectly classified instances (ICI), accuracy and recall metrics as shown in Table 4.3. These metrics depend on true positive (a), false positive (b), true negative(c) and false negative (d) as follow:

True Positive (a): Number of P2P class that are correctly classified.
False Positive (b): Number of P2P that are classified as the non-P2P.
True Negative (c): Number of non-P2P (nP2P) class that are correctly classified.
False Negative (d): Number of nP2P class that are classified as P2P class.

TABLE 4.3: Verification criteria

Criterion	Symbol	Expression
incorrectly classified instances	ICI	b+d
Recall	Recall	a/(a+d)
Accuracy	Accuracy	(a+d)/(a+b+c+d)

5. Results and Discussion

This subsection explains the feature selection results. Firstly, we apply six algorithms of features selection to generate candidate subset of features from online features without IAT. The features are selected for each training dataset using Chi-squared, IG, GR, FRSE, FAE and PCA algorithms. Secondly, we evaluate the selected features presented in Table 5.1 using six ML algorithms NaiveBayes, J48, Lib SVM, HoeffdingTree and SMO. Thirdly, we select the best feature subset that is suitable for online P2P traffic detection by examining the combined subset of features. Fourthly, we study the impact of the integration of online features and inter-arrival time. Lastly, the viability of the proposed approach in Table 5.3.

© IEOM Society International

5.1 Feature selection methods results

This subsection explains generated candidate subset of features from online features without IAT using six features selection algorithms as shown in Table 5.1.

Table 5.1 The candidate feature subset using feature selection algorithms (see Table 1)

FS method	Features
Chi	1,2,8,9,7,10,15
GR	2,4,5,6,7,9,8,10,15
IG	2,1,10,8,9,7,5,15
PCA	1,2,4,5,6,7,15
Fuzzy Rough Sub set Eval	1,2,8,9,10,11,12,15
Filtered Attribute Eval	1,2,10,8,7,9,5,15

5.2 CLASSIFICATION RESULTS

The performance results of selected feature selection methods for machine learning classifiers are discussed here. Figure 5.1 shows accuracy results for different feature selection techniques. The result shows that the combination of PCA feature selection algorithm with most classifiers algorithm is the best combination in terms of accuracy, consistently the combination of PCA feature selection algorithm and J48 out performance of all classifiers in terms of accuracy with an accuracy of (99.799%), while the combination of GR feature selection algorithm with SMO gave low accuracy.

Figures 5.2 shows the results of ICI, the result shows that the combination of PCA feature selection algorithm with most classifiers algorithm is the best combination in term of ICI, consistently the combination of PCA feature selection algorithm and J48 out performance of all classifiers in term of ICI is (0.2001%), while the combination of GR feature selection algorithm with SMO gave a high ICI.

Figures 5.3 shows the results of recall, the result shows that the combination of PCA feature selection algorithm with most classifiers algorithm is the best combination in term of recall, consistently the combination of PCA feature selection algorithm and J48 out performance of all classifiers in term of recall is (0.998) while the combination of GR feature selection algorithm and SMO gave a low recall.

Figures 5.4 shows the results for testing time, the result shows that the combination of PCA feature selection algorithm with all classifiers algorithm is the best combination in term of testing time, consistently the combination of PCA feature selection algorithm and NaiveBayes out performance of all classifiers in term of testing time it is (0.002/s), while the combination of IG feature selection algorithm with Lib_SVM gave a high testing time.

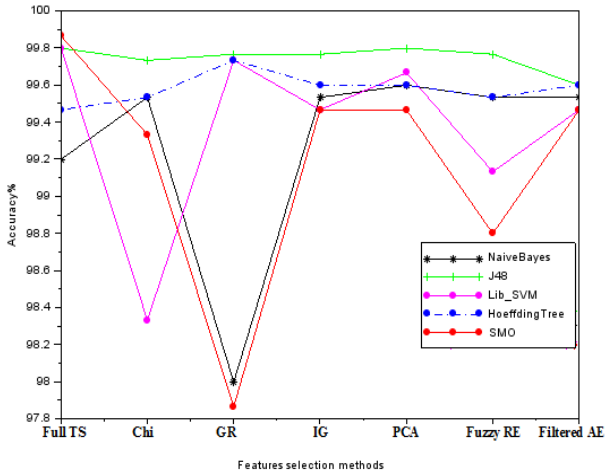

Figure 5.1 Accuracy of different feature selection methods with respect to five classifiers

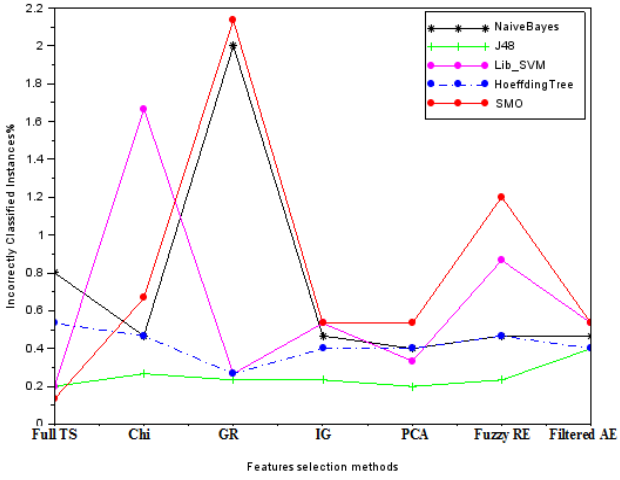

Figure 5.2 Incorrectly classified instances (ICI) of different feature selection methods with respect to five classifiers

© IEOM Society International

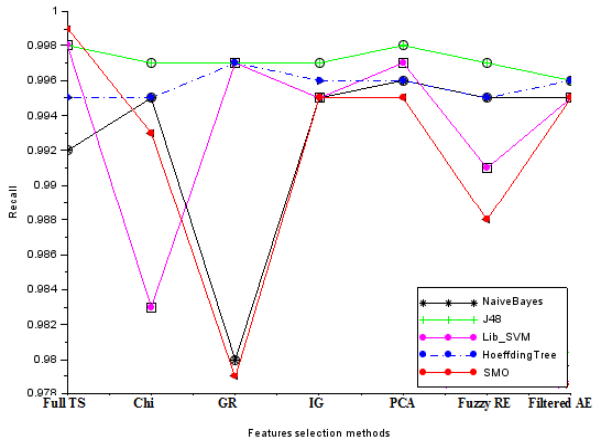

Figure 5.3 Recall of different feature selection methods with respect to five classifiers

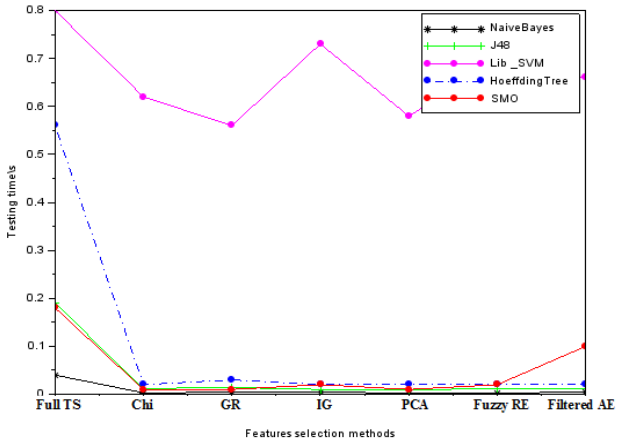

Figure 5.4 Testing time of different feature selection methods with respect to five classifiers.

Table 5.2 shows the proposed feature subset for online P2P internet traffic detection.

No	Abbreviation	Description
1	Destination port number	DP
2	Source port number	SP
4	Bytes in Ethernet packet	B_Eth
5	Bytes in IP packet	B_IP
6	Control bytes in packet	C_bp
7	Bytes in Ethernet packet (uplink)	UB_Eth
15	Protocol	Application class

5.3 Impact of the integration of online features and IAT

The impact of the integration of online features and inter-arrival time are discussed here. Figures 5.5 - 5.7 show the integration of online features and inter-arrival time (IRT) decreases accuracy and recall, and also increase incorrectly classified percentage instances. These results because IAT morphing usually involves alternation on direction pattern and depend on different network locations. These dependence on integration with other features. Figure 5.8 shows the integration of online features and inter-arrival time (IRT) increase testing time. This due to the inter-arrival time (IRT) needs pre-processing such as normalization in order to become significant. In this experiment using J48 classifier algorithm as gives the best effectiveness (accuracy, ICI and recall) and Naïve Bayes algorithm gives the best efficiency (testing time).

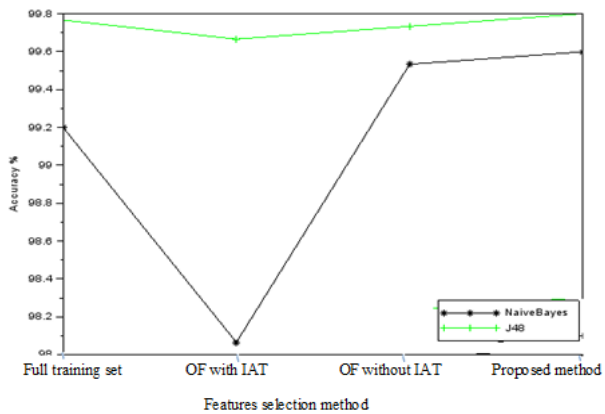

Figure 5.5 The impact of the integration of online features and inter-arrival time in accuracy.

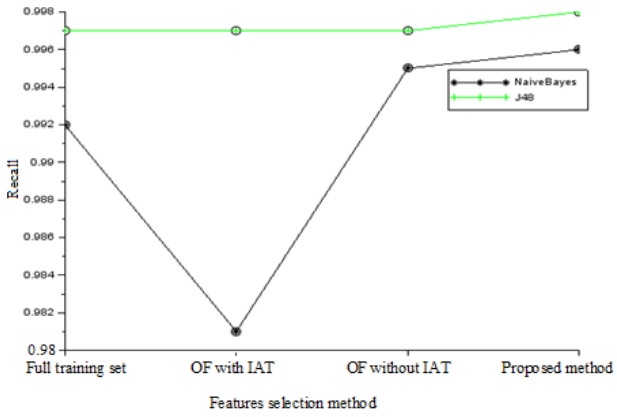

Figure 5.6 The impact of the integration of online features and inter-arrival time in recall

© IEOM Society International

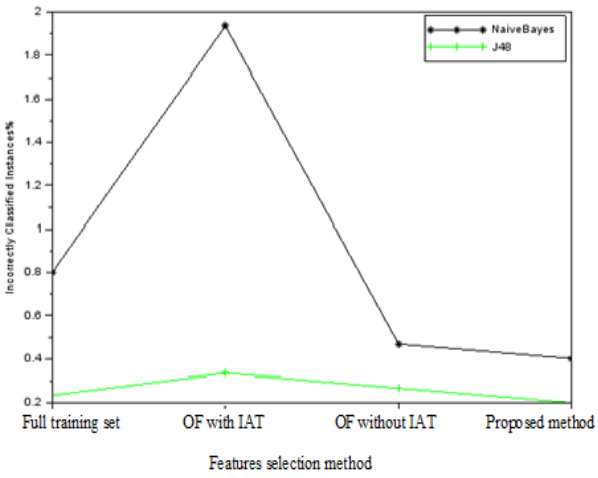

Figure 5.7 The impact of the integration of online features and inter-arrival time in ICI

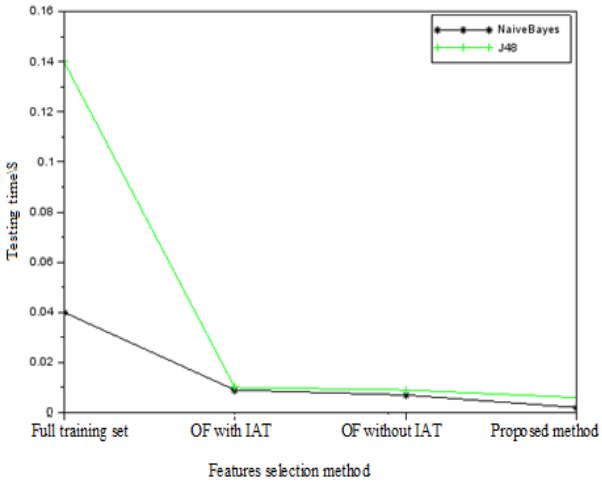

Figure 5.8 The impact of the integration of online features and inter-arrival time in testing time.

Table 5.3 shows the results comparison between the proposed approach and work proposed in [7, 8]. The works in [7, 8] used ten Cambridge datasets and Naïve Bayes to evaluate the feature selection algorithms named (BFS) and SOF in order. Our proposed method improves accuracy up to 8.57% and 1.79% in order for Naïve Bayes, whilst, shows the smallest testing time compared with works proposed in [7, 8]. This improvement is results of reducing the number of features for our system as well as all selected features can be applied online.

6. Conclusion

Classification of bandwidth heavy Internet traffic is important for network administrators to throttle the network from heavy bandwidth traffic applications. Statistical methods have been proposed as promising method to identify internet traffic based on their statistical features. The selection of statistical features plays an important role in the classification results. In this paper, we proposed a subset of features for online P2P traffic classification. The experimental results indicate that our proposed online features result in an improved accuracy and uses smaller detection time.

Table 5.3 Effectiveness of the proposed approach

	Work in [22] Naive Bayes BFS	Work in [8] Naive Bayes SOF	Proposed Naive Bayes	Proposed J48
Accuracy	90.92(%)	97.80 (%)	99.59 (%)	99.799%
Recall	0.60	0.97	0.99	0.999
Number of features	11	8	7	7

References

[1] Johnson, D.L., Belding, E.M., and Van Stam, G. (2012). Network traffic locality in a rural African village. In Proceedings of the fifth international conference on information and communication technologies and development. 2012 ACM New York. Atlanta, USA. 12-15 March 2012. 268-277.

[2] Torres, R.D., Hajjat, M.Y., Rao, S.G., Mellia, M., and Munafò, M.M. (2009). Inferring undesirable behavior from P2P traffic analysis. ACM SIGMETRICS Performance Evaluation Review, 2009 ACM New York. NY, USA. 3 December 2009 . 25-36.

[3] Sen, S., Spatscheck, O., and Wang, D. (2004). Accurate, scalable in-network identification of p2p traffic using application signatures. 13th international conference on World Wide Web, 2004 ACM New York. NY, USA. 17 - 22 May 2004. 512-521.

[4] Moore, A.W., and Papagiannaki, K. (2005). Toward the accurate identification of network applications. In Passive and Active Network Measurement. Springer, Vol. 3431, 41-54.

[5] Van Der Putten, P., and Van Someren, M. (2004). A bias-variance analysis of a real world learning problem: The CoIL challenge 2000. Machine Learning,vol. 57, 177-195.

[6] ZHAO, J.-j., HUANG, X.-h., Qiong, S., and Yan, M. (2008). Real-time feature selection in traffic classification. The Journal of China Universities of Posts and Telecommunications, vol. 15, 68-72.

© IEOM Society International

[7] Zhang, H., Lu, G., Qassrawi, M.T., Zhang, Y., and Yu, X. (2012). Feature selection for optimizing traffic classification. Computer Communications,vol. 35, 1457-1471.
[8] Jamil, H.A., Mohammed, A., Hamza, A., Nor, S.M., and Marsono, M.N. (2014). Selection of on-line features for peer-to-peer network traffic classification. In Recent Advances in Intelligent Informatics. Springer.
[9] Nguyen, T.T., and Armitage, G. (2008). A survey of techniques for internet traffic classification using machine learning. Communications Surveys & Tutorials, IEEE, vol. 10, 56-76.
[10] Erman, J., Mahanti, A., Arlitt, M., Cohen, I., and Williamson, C. (2007). Offline/realtime traffic classification using semi-supervised learning. Performance Evaluation, vol. 64, 1194-1213.
[11] Ma, Y., Qian, Z., Shou, G., and Hu, Y. (2008). Study of information network traffic identification based on C4. 5 algorithm. In Wireless Communications, Networking and Mobile Computing, 2008 WiCOM'08 4th International Conference. IEEE, China 12-14 Oct 2008.1-5.
[12] Schultz, M.G., Eskin, E., Zadok, F., and Stolfo, S.J. (2001). Data mining methods for detection of new malicious executables. Security and Privacy, 2001. S&P 2001. 2001 IEEE Symposium.Oakland, CA.14-16 May 2000. 38-49
[13] Tahan, G., Rokach, L., and Shahar, Y. (2012). Mal-ID: Automatic Malware Detection Using Common Segment Analysis and Meta-Features. The Journal of Machine Learning Research, vol. 98888, 949-979.
[14] Moore, A., Zuev, D., and Crogan, M. (2005). Discriminators for use in flow-based classification. Technical report, Intel Research, Cambridge.
[15] Auld, T., Moore, A.W., and Gull, S.F. (2007). Bayesian neural networks for internet traffic classification. Neural Networks, IEEE Transactions ,vol.18, 223-239.
[16] Moore, A.W., and Zuev, D. (2005). Internet traffic classification using bayesian analysis techniques. In ACM SIGMETRICS Performance Evaluation Review. Number 1, June 2005. 50-60.
[17] Jun, L., Shunyi, Z., Shidong, L., and Ye, X. (2007). P2P traffic identification technique. In Computational Intelligence and Security, 2007 International Conference on IEEE. Heilongjiang .15-19 Dec 2007. 37-41.
[18] Yang, Y.-x., Wang, R., Liu, Y., and Zhou, X.-y. (2007). Solving p2p traffic identification problems via optimized support vector machines. Computer Systems and Applications, 2007 AICCSA'07 IEEE/ACS International Conference. Amman.13-16 May 2007.165-171.
[19] Monemi, A., Zarei, R., and Marsono, M.N. (2013). Online NetFPGA decision tree statistical traffic classifier. Computer Communications, vol. 36, 1329-1340.
[20] Qu, B., Zhang, Z., Zhu, X., and Meng, D. (2015). An empirical study of morphing on behavior‐based network traffic classification. Security and Communication Networks, vol. 8, 68-79.
[21] Kögel, J. (2013). One-Way Delay Measurement based on Flow Data in Large Enterprise Networks.Univ. Stuttgart, Inst. für Kommunikationsnetze und Rechnersysteme.
[22] Zhen, L., and Qiong, L. (2012). A new feature selection method for internet traffic classification using ml. Physics Procedia, vol. 33, 1338-1345.
[23] Bernaille, L., Teixeira, R., and Salamatian, K. (2006). Early application identification. CoNEXT 2006, Lisboa, Portugal. 4-7 December 2006, 6.
[24] Liu, H., and Setiono, R. (1995). Chi2: Feature selection and discretization of numeric attributes. In tai. Herndon, 5-8 Nov 1995, 388.
[25] Moskovitch, R., Stopel, D., Feher, C., Nissim, N., Japkowicz, N., and Elovici, Y. (2009). Unknown malcode detection and the imbalance problem. Journal in Computer Virology, vol. 5, 295-308.
[26] Henchiri, O., and Japkowicz, N. (2006). A feature selection and evaluation scheme for computer virus detection. Sixth International Conference on Data Mining (ICDM'06). Hong Kong .18-22 Dec 2006.891-895.
[27] Radzikowska, A.M., and Kerre, E.E. (2002). A comparative study of fuzzy rough sets. Fuzzy sets and systems, vol. 126, 137-155.
[28] Wang, W., Zhang, X., and Gombault, S. (2009). Constructing attribute weights from computer audit data for effective intrusion detection. Journal of Systems and Software, vol. 82, 1974-1981.
[29] Wang, W., Guan, X., and Zhang, X. (2008). Processing of massive audit data streams for real-time anomaly intrusion detection. Computer Communications, vol. 31, 58-72.
[30] Lei, D., Xiaochun, Y., and Jun, X. (2008). Optimizing traffic classification using hybrid feature selection. In Web-Age Information Management, 2008 WAIM'08 The Ninth International Conference on.IEEE. Zhangjiajie, China . 20-22 July 2008. 520-525.
[31] Tan, P.-N., Steinbach, M., and Kumar, V. (2006). Introduction to data mining.Pearson Addison Wesley Boston.
[32] Committee, I.C.S.L.M.S. (1997). Wireless LAN medium access control (MAC) and physical layer (PHY) specifications. In. IEEE Std.
[33] Gringoli, F., Salgarelli, L., Dusi, M., Cascarano, N., and Risso, F. (2009). Gt: picking up the truth from the ground for internet traffic. ACM SIGCOMM Computer Communication Review, vol. 39, 12-18.

© IEOM Society International